WHAT IS CLAIMED IS:

1. A method of determining a hybrid text summary comprising the steps of:

determining discourse constituents for a text;

determining a structural representation of discourse for the text;

determining relevance scores for discourse constituents based on at least one non-structural measure of relevance;

percolating relevance scores based on the structural representation of discourse;

determining a hybrid text summary based on discourse constituents with relevance scores compared to a threshold relevance score.

2. The method of claim 1, wherein the structural representation of discourse is determined based on a theory of discourse analysis.

3. The method of claim 2, wherein the theory of discourse analysis is at least one of: the Linguistic Discourse Model, the Unified Linguistic Discourse Model, Rhetorical Structure Theory, Discourse Structure Theory and Structured Discourse Representation Theory.

4. The method of claim 1, wherein non-structural measures of relevance are determined based on at least one of: statistics, keywords, knowledge bases.

5. The method of claim 1, wherein percolating the relevance scores comprises the steps of:

for each child discourse constituent node in the structural representation, assigning the relevance score of the child discourse constituent node to the parent discourse constituent node if the child discourse constituent node is more relevant;

for any subordinating nodes, assigning the relevance scores of the subordinated discourse constituent to the subordinating discourse constituent if the subordinated discourse constituent is more relevant; and

for any coordination nodes, assigning the relevance score of the most relevant child to other child discourse constituent nodes.

6. The method of claim 1, wherein percolating the relevance scores comprises the steps of:

for each child discourse constituent node in the structural representation, assigning the relevance score of the child discourse constituent node to the parent discourse constituent node if the child discourse constituent node is more relevant than its parent;

for each coordinated discourse constituent node, assigning the relevance score of the coordinated discourse constituent node to each preceding less relevant sibling node;

for each child discourse constituent node that is not a coordinated discourse constituent node and is not a subordinated discourse constituent node, assigning the relevance score of the parent discourse constituent node to the child discourse constituent node if the parent discourse constituent is more relevant than the child;

for each coordinated discourse constituent node, assigning the relevance score of the parent discourse constituent node to the coordinated discourse constituent node, if the coordinated discourse node and all its siblings are less relevant than the parent node;

for each subordinated discourse constituent node, assigning the relevance score of the subordinated discourse constituent node to the subordinating discourse constituent if the subordinated discourse constituent is more relevant than the subordinating node; and

for each node, repeating these steps, until no node can be found whose relevance score is changed to the relevance score of another node.

7. The method of claim 6, wherein the percolation of relevance scores is applied to progressively larger sets of linked nodes.

8. A method of determining a hybrid text summary comprising the steps of:

determining discourse constituents for a text;

determining a structural representation of discourse for the text;

determining relevance scores for discourse constituents;

percolating relevance scores based on the structural representation of discourse comprising the steps of:

for each discourse constituent leaf node, determining the number of subordinated edges plus one;

determining a score based on the inverse of the number of subordinated edges +1;

for each discourse constituent node, assigning the score of a child discourse constituent node to the parent discourse constituent node, if the score is less relevant;

for any subordination discourse constituent node, assigning the score of the subordinated discourse constituent node to the subordinating discourse constituent node if the subordinated discourse constituent score is lower;

assigning the relevance scores of any coordination discourse constituent node to each child discourse constituent of the coordination if it is lower;

determining an adjusted relevance score based on the score and the subordination level; and

determining a hybrid text summary based on discourse constituents with relevance scores compared to a threshold relevance score.

9. The method of claim 1, further comprising the steps of:

determining every leaf discourse constituent containing an anaphor;

for each anaphor, determine any unique antecedent** referents for the anaphor;

substituting the unique antecedent referent into the leaf discourse constituent for

the anaphor;

removing the discourse constituent containing the unique antecedent referent from the set of the discourse constituents with relevance scores more relevant that the threshold relevance score.

10. The method of claim 1, wherein percolation of relevance scores comprises the steps of:

determining important discourse constituent nodes;

determining unresolved anaphors;

determining potential resolving discourse constituents **containing potential antecedent referent with potential to resolve anaphors;

percolating relevance score of important discourse constituents through a reduced span of potential resolving discourse constituents; and

determining a reduced span of discourse constituents based on relevance score.

11. A system for determining hybrid text summaries comprising:

an input/output circuit for retrieving a text;

a processor for determining discourse constituents for the text and attaching the discourse constituents into a structural representation of discourse;

a relevance score determination circuit for determining relevance scores for the discourse constituents based on at least one non-structural measure of relevance;

a percolation circuit for percolating discourse constituent relevance scores based on the structural representation of discourse; and where the processor determines a hybrid text summary based on the discourse constituents with relevance scores exceeding a threshold relevance score.

12. The system of claim 11, wherein the structural representation of discourse is determined based on a theory of discourse analysis.

13. The system of claim 12, wherein the theory of discourse analysis is at least one of: the Linguistic Discourse Model, the Unified Linguistic Discourse Model, Rhetorical Structure Theory, Discourse Structure Theory and Structured Discourse Representation Theory.

14. The system of claim 11, wherein non-structural measures of relevance are determined based on at least one of: statistics, keywords, knowledge bases.

15. The system of claim 11, wherein, for each child discourse constituent node in the structural representation, the percolation circuit assigns the relevance score of the child discourse constituent node to the parent discourse constituent node if the child discourse constituent node is more relevant;

for any subordinating nodes, the percolation circuit assigns the relevance scores of the subordinated discourse constituent to the subordinating discourse constituent if the subordinated discourse constituent is more relevant; and

for any coordination nodes, the percolation circuit assigns the relevance score of the most relevant child to other child discourse constituent nodes.

16. The system of claim 11, wherein for each child discourse constituent node in the structural representation, the percolation circuit assigns the relevance score of the child discourse constituent node to the parent discourse constituent node if the child discourse constituent node is more relevant than its parent;

for each coordinated discourse constituent node, the percolation circuit assigns the relevance score of the coordinated discourse constituent node to each preceding less relevant sibling node;

for each child discourse constituent node that is not a coordinated discourse constituent node and is not a subordinated discourse constituent node, the percolation circuit assigns the relevance score of the parent discourse constituent node to the child discourse constituent node if the parent discourse constituent is more relevant than the child;

for each coordinated discourse constituent node, the percolation circuit assigns the relevance score of the parent discourse constituent node to the coordinated discourse constituent node, if the coordinated discourse node and all its siblings are less relevant than the parent node;

for each subordinated discourse constituent node, the percolation circuit assigns the relevance score of the subordinated discourse constituent node to the subordinating discourse constituent if the subordinated discourse constituent is more relevant than the subordinating node; and

for each node, repeating these steps, until the percolation circuit can find no node whose relevance score is changed to the relevance score of another node.

17. The system of claim 16, wherein the percolation is applied to progressively larger sets of linked nodes.

18. A system for determining hybrid text summaries comprising:

an input/output circuit for retrieving a text;

a processor for determining discourse constituents for the text and attaching the discourse constituents into a structural representation of discourse;

a relevance score determination circuit for determining relevance scores for the discourse constituents based on at least one non-structural measure of relevance;

a percolation circuit for percolating discourse constituent relevance scores based on the structural representation of discourse; wherein for each discourse constituent leaf node, the percolation circuit determines the number of subordinated edges plus one and an score based on the inverse of the number of subordinated edges +1;

for each discourse constituent node, the percolation circuit assigns the score of a child discourse constituent node to the parent discourse constituent, if the score is less relevant;

for any subordination discourse constituent node, the percolation circuit assigns the score of the subordinated discourse constituent node to the subordinating discourse constituent node if the subordinated discourse constituent score is lower;

the percolation circuit assigns the scores of any coordination discourse constituent node to each child discourse constituent of the coordination if it is lower; and the processor determines an adjusted relevance score based on the score and the subordination level; and a hybrid text summary based on the discourse constituents with relevance scores exceeding a threshold relevance score.

19. The system of claim 11, wherein the processor determines every leaf discourse constituent containing an anaphor;

for each anaphor, the processor determines any unique preceding referents for the anaphor;

the processor substitutes the unique antecedent referent into the preceding

discourse constituent for the anaphor referent; and

the processor removes the preceding discourse containing the unique referent from the discourse constituents with relevance scores exceeding the threshold relevance score.

20. The system of claim 11, the percolation circuit determines every leaf discourse constituent containing an anaphor;

for each anaphor, the percolation circuit determines any unique preceding referents for the anaphor;

the percolation circuit substitutes the unique antecedent referent into the leaf discourse constituent for the anaphor;

the percolation circuit removes the discourse constituent containing the unique antecedent referent from the set of the discourse constituents with more relevant relevance scores.

21. The system of claim 11, wherein the processor determines important discourse constituent nodes based on a non-structural measure of relevance;

determines unresolved referents in the important discourse constituents;

determines potential resolving discourse constituents with potential to resolve referents;

percolates relevance score of important discourse constituents through a reduced span of potential resolving discourse constituents; and

determines a reduced span of discourse constituents based on relevance score.

22. A carrier wave encoded to transmit a control program, useable to program a computer to determine hybrid text summary, to a device for executing the program, the control program comprising:

instructions for determining discourse constituents for a text;

instructions for determining a structural representation of discourse for the text;

instructions for determining relevance scores for discourse constituents based on at least one non-structural measure of relevance;

instructions for percolating relevance scores based on the structural representation of discourse;

instructions for determining a hybrid text summary based on discourse constituents with relevance scores compared to a threshold relevance score.

23. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code usable to program a computer to determine hybrid text summary comprising the steps of:

determining discourse constituents for a text;

determining a structural representation of discourse for the text;

determining relevance scores for discourse constituents based on at least one non-structural measure of relevance;

percolating relevance scores based on the structural representation of discourse;

determining a hybrid text summary based on discourse constituents with relevance scores compared to a threshold relevance score.

24. A method for discourse parsing comprising the steps of:

determining a structural representation of discourse based on a theory of discourse analysis;

determining at least one sentence of a text;

determining sentential-level parse features for the at least one sentence;

determining a mapping between the sentential-level parse features and discourse-level parse features;

determining a discourse-level parse tree of the at least one sentence based on the mapping;

determining a main discourse constituent for the at least one sentence;

determining an attachment of the discourse level parse tree to the structural representation of discourse by the determined main discourse constituent based on attachment rules for the theory of discourse.

25. A method of segmenting text into discourse constituents comprising the steps of:

determining a theory of discourse analysis;

determining candidate segments;

determining attributes of candidate segments associated with continuing the discourse;

determining if the candidate segment is a discourse constituent based on the theory of discourse analysis and the determined attributes.

26. The method of claim 25, wherein the attributes are determined based on at least one of: a part-of-speech tag, a probabilistic parser, a statistical parser, a finite state parser, a symbolic parser, a lexicon, and a WordNet relation.

27. A method of determining a structural representation of discourse

comprising the steps of:

determine discourse constituents for a text; and

conjoining the discourse constituents into a structural representation of discourse based on theory of discourse analysis classifications of the discourse constituents and at least one of a syntactic, a semantic and a lexical-semantic constraint.

28. The method of determining a hybrid text summary for a text comprising the method of segmenting text of claim 24, the method of determining a structural representation of discourse of claim 27, the discourse parsing of claim 23 and the method of determining a hybrid text summary of claim 1.

29. A system for discourse parsing comprising:

an input/output circuit;

a processor which determines a structural representation of discourse based on a theory of discourse analysis;

determines at least one sentence of a text;

determines sentential-level parse features for the at least one sentence;

determines a mapping between the sentential-level parse features and discourse-level parse features;

determines a discourse-level parse tree of the at least one sentence based on the mapping;

determines a main discourse constituent for the at least one sentence;

determining an attachment of the discourse level parse tree to the structural representation of discourse by the determined main discourse constituent based on attachment rules for the theory of discourse.

30. A system for segmenting text into discourse constituents comprising:

an input/output circuit;

a processor which determines a theory of discourse analysis;

determines candidate segments;

determines attributes of candidate segments associated with continuing the discourse;

determines if the candidate segment is a discourse constituent based on the theory of discourse analysis and the determined attributes.

31. The system of claim 30, wherein the attributes are determined based on

at least one of: a part-of-speech tag, a probabilistic parser, a statistical parser, a finite state parser, a symbolic parser, a lexicon, and a WordNet relation.

32. A system of determining a structural representation of discourse comprising:

an input/output circuit;

a processor which determines discourse constituents for a text; and conjoins the discourse constituents into a structural representation of discourse based on theory of discourse analysis classifications of the discourse constituents, and at least one of a syntactic, a semantic and a lexical-semantic constraint.

33. The hybrid text summarization system comprising the system for segmenting text of claim 30, the method of determining a structural representation of discourse of claim 32, the discourse parsing of claim 29 and the method of determining a hybrid text summary of claim 11.

34. The method of claim 8, further comprising the steps of:

determining a combined relevance score based on the relevance scores and non-structural relevance scores and percolating the combined relevance score.

35. The system of claim 18, the relevance circuit determines combined relevance score based on the relevance scores and non-structural relevance scores and percolating the combined relevance score.

36. A hybrid text summarization system comprising:

means for determining discourse constituents for a text;

means for determining a structural representation of discourse for the text;

means for determining relevance scores for discourse constituents based on at least one non-structural measure of relevance;

means for percolating relevance scores based on the structural representation of discourse; and

means for determining a hybrid text summary based on discourse constituents with relevance scores compared to a threshold relevance score.

37. A hybrid text summarization system comprising:

means for determining discourse constituents for a text;

means for determining a structural representation of discourse for the text;

means for determining relevance scores for discourse constituents;

means for percolating relevance scores based on the structural representation of discourse comprising the steps of:

means for for each discourse constituent leaf node, determining the number of subordinated edges plus one;

means for determining a score based on the inverse of the number of subordinated edges +1;

means for for each discourse constituent node, assigning the score of a child discourse constituent node to the parent discourse constituent node, if the score is less relevant;

means for for any subordination discourse constituent node, assigning the score of the subordinated discourse constituent node to the subordinating discourse constituent node if the subordinated discourse constituent score is lower;

means for assigning the relevance scores of any coordination discourse constituent node to each child discourse constituent of the coordination if it is lower;

means for determining an adjusted relevance score based on the score and the subordination level; and

determining a hybrid text summary based on discourse constituents with relevance scores compared to a threshold relevance score.

38. A method for discourse parsing system comprising:

means for determining a structural representation of discourse based on a theory of discourse analysis;

means for determining at least one sentence of a text;

means for determining sentential-level parse features for the at least one sentence;

means for determining a mapping between the sentential-level parse features and discourse-level parse features;

means for determining a discourse-level parse tree of the at least one sentence based on the mapping;

means for determining a main discourse constituent for the at least one sentence; and

means for determining an attachment of the discourse level parse tree to the structural representation of discourse by the determined main discourse constituent based on attachment rules for the theory of discourse.

39. A text segmenting system comprising:

means for determining a theory of discourse analysis;

means for determining candidate segments;

means for determining attributes of candidate segments associated with continuing the discourse; and

means for determining if the candidate segment is a discourse constituent based on the theory of discourse analysis and the determined attributes.